

EVALUATION AND EFFICIENCY OF RESEARCH AND EXPERIMENTAL DEVELOPMENT (R&D) INSTITUTIONS AND UNITS

Simon Schwartzman and Amaury de Souza

March 1984

1 - The question of R&D evaluation

Evaluation is a crucial component of any R&D activity as well as one of the most difficult tasks to handle. Whether basic research or experimental development is concerned, scientific work is expected to be performed at the highest possible level of competence. R&D efforts are justified by the expectation that it will reach beyond what is already known and established, opening up new horizons for knowledge, creating novel and better ways of dealing with the world and using it for the benefit of mankind. To work at the frontier of knowledge, making the best use of available technical and intellectual resources, is part of the very definition of R&D.

Scientific research and development activities, however, are not easily amenable to objective evaluation. In the first place, working at the frontier of knowledge, scientists can rightly claim that it is upon them to evaluate their own work, or, at least, that no such evaluation can be carried out without their active participation. It follows that scientific research entails a necessary element of self-evaluation and control, which may lead in turn to difficult problems of professional and institutional self-protection and deception¹.

Second, scientific research and development are usually open-ended activities, in the sense that it is difficult or even impossible to determine in advance their final product. There are, of course, significant differences between more academic or "pure" and more applied fields of research. However, even when precise specifications of final products do

¹ The notion that the scientific enterprise is strongly tainted by corporate self-protection and deception appears in William Broad and Nicolas Wade, *Betrayers of the Truth: Fraud and Deceit in the Halls of Science* (New York: Simon and Schuster, 1982). For a rebuttal, see Henry A. Bauer, "Betrayers of the Truth: A Fraudulent and Deceitful Title from the Journalists of Science", *4s Review*. Vol.1, N.3, Fall 1983

exist, it is normally impossible to fully anticipate the steps required to bring them about, since such steps constitute an integral part of the research effort.

Finally, there is hardly any consensus as to the specific meaning of "performance" and quality. Thus, the problem of scientific research evaluation is not only that there are no clear and indisputable judgmental parameters and procedures but also that scientists, policymakers and relevant public have different and often contradictory expectations regarding the outcomes of scientific work as well as the standards of excellence by which it should be evaluated.

2 - R&D evaluation in developing countries

Research and development evaluation in countries that do not have a firmly established science and technology tradition poses further difficulties. It remains a matter of scholarly dispute that were the precise mechanisms that led countries such as Germany, England, France and the United States to play a central role in the development of modern science and technology. It is generally accepted, however, that the emergence of advanced scientific centers was historically linked with the existence of strong and fairly autonomous universities, intensive industrialization and the diffusion of social values that stressed the centrality of reason and empirical knowledge and placed a premium on individual effort².

Government support and sponsorship of scientific and technological research was another important component of past national experiences, but not necessarily the major one. In modern times, however, there is a clear parallel between the roles played by governments in science and technology and in economic development. Latecomers to the industrial world display higher levels of state participation in the economy than first comers³. The same is true of countries that developed scientific and technological capabilities in late nineteenth and early twentieth centuries, such as Japan, the Soviet Union, and India.

² The conditions for the emergence of modern science are discussed in the seminal works of Robert K. Merton, *Science, Technology and Society in Seventeenth -Century England* (New York: Harper & Row, 1970) and Joseph Ben-David, *The Scientists' Role in Society: a Comparative Study* (Englewood Cliffs, NJ: Prentice-Hall, 1971), pp.17-23.

³ See A. Gerschenkron, *Economic Backwardness in Historical Perspective* (Cambridge, Massachusetts: Harvard University Press, 1962).

After World War II, the notion that science and technology could not exist and develop without government support and guidance became widely accepted. The war effort demonstrated that scientific and technological advances were key ingredients of economic and military power and economic theories were put forward stressing the role of education and technological change as crucial components of economic growth. However, as science and technology expanded the radius of its potential contribution to societal welfare governments in most countries were hard pressed to increase investments in R&D activities and higher education. Government involvement in science and technology was also furthered by the growing awareness that public scrutiny and control was needed to prevent the misuse of scientific knowledge and technical capabilities.

The increasing role of government in science and technology also raised questions of efficiency that had been traditionally associated with the production of goods and services. In a free market it is expected that open competition will naturally drive out the less efficient and competent⁴. Sociologists of science have used this market analogy to explain how the scientific community is organized, and how resources and awards of prestige and authority are distributed through mechanisms of competition and self-regulation⁵. Carrying the analogy further, however, leads to the recognition that scientific research and development are also liable to monopoly or oligopoly growth and stagnation. In addition, the existence of an international science and technology Market demands purposeful intervention to reverse or compensate for the ensuing geographic concentration of talent and resources.

If governmental intervention seeks to redress imbalances produced by open competition, then standards of evaluation derived from the operation of the free market cannot be readily applied to protected scientific and technological activities. Indeed, unhampered competition between national R&D systems unequally endowed with scientific and

⁴ For the importance of market mechanisms and alternative ways for fostering efficiency, see A. O. Hirschman, *Exit, Voice and Loyalty - Response to Decline in Firms, Organizations and States* (Cambridge, Massachusetts: Harvard University Press, 1970).

⁵ The classic reference is M. Polanyi, *Personal Knowledge* (Chicago: Chicago University Press, 1958). See also Robert K. Merton, "The Normative Structure of Science", in R. K. Merton, *The Sociology of Science* (Chicago: The University of Chicago Press, 1973).

technological capabilities is likely to entail the destruction of incipient local efforts toward scientific and industrial growth and its potential utilization to the benefit of developing societies. At the same time, protection against market competition is often transformed into complacency toward incompetence, irresponsibility, and waste. Unproductive research undertakings may drain public resources for long periods and barriers to technology transfers or the closure of markets to foreign products may substitute the protection of industrial inefficiency for the original goal of stimulating development.

It is possible to argue that the creation and development of modern science and technology institutions and policymaking bodies was the paramount concern of developing countries in the 1960s and 1970s and that considerations of effectiveness were accordingly granted low priority. Not infrequently, R&D evaluation was seen as premature, hence harmful to ambitious programs of institution building and scientific development. An opposite movement may be in the offing in the 1980s. Faith in the power of governmental planning and intervention has decreased⁶. While science policymaking agencies are now to be found everywhere, there are persistent doubts as to whether their cost and the complexity they bring to bear over scientific and technological activities are justified by tangible results. Furthermore, the number of research and experimental development institutions and units has greatly increased in many developing countries, but their average productivity ranks low by most standards, and there is mounting evidence to the effect that they can hardly be expected to improve just through more generous funding or technical assistance⁷.

The uncertainty that now surrounds R&D efforts in developing countries is likely to affect the volume of resources devoted to science and technology although their levels of R&D expenditures still rank far below the developed world's. Viewed in this context, sustained support for scientific and technological development will be hard to come by in

⁶ See Naomi Caiden and Aaron Wildawsky, *Planning and Budgeting in Poor Countries* (New York: John Wiley & Sons, 1974).

⁷ On the quality of third world research, see Eugene C. Garfield, "Mapping Science in the Third World", *Science and Public Policy*. June 1973. pp. 112-127.

developing countries unless current efforts are properly evaluated and reoriented.

3 - Conflicting criteria for R&D evaluation

Scientific research and experimental development include a plurality of activities and objectives that can hardly be subjected to the same judgmental criteria. Empirical studies show that research units and institutions hold different objectives and orientations, making it impossible to gauge their effectiveness in terms of a single criterion⁸. Scientific effectiveness entails at least the following dimensions⁹:

- Academic effectiveness, or the contribution toward the development of scientific knowledge, as indicated by scientific publications, participation in professional meetings and the like;
- Technological effectiveness, or the ability to bring about new experimental devices and products;
- Training effectiveness, or the capacity to form new researchers and provide technical and scientific training;
- Social effectiveness, or the contribution toward the solution of socially pressing problems; and
- Economic effectiveness, or the ability to make products with higher market value than their R&D cost.

The fact that dimensions of R&D effectiveness are uncorrelated, as several empirical studies have concluded, means that different research institutions and units emphasize different goals and that an outstanding performance with respect to any one objective is no indication of similar performance on all others. It is indeed a common mistake to elect

⁸ "In most general terms, it can be said that efficiency is a concept intrinsic to Science and Technology which measures how far resources invested in R&D have been productive within reasonable time limits; it can theoretically be judged by input - output ratios whereas effectiveness is a concept extrinsic to S&T which gauges the output of RID both qualitatively and quantitatively against socioeconomic goals or objectives pursued." See Unesco, *Science. Technology and Governmental Policy - a Ministerial Conference for Europe and North America*. Minespol II, 1979. p. 42.

⁹ The multidimensionality of research performance has been one of the most consistent results of the "International Comparative Study on the Research Performance of Research Units" which is being conducted worldwide under the coordination of Unesco. The following classification is adapted from Frank Andrews (editor). *Scientific Productivity* (Cambridge University Press and UNESCO. 1979, p. 39).

one single dimension of effectiveness as paramount and then use it as the sole standard for evaluation purposes. Planners will often demand economic or social effectiveness while dismissing the quest for academic or technical excellence as self-serving and wasteful; academics will argue that the production of new knowledge is the precondition of all scientific and experimental work; and university authorities may insist on the evaluation of research groups in terms of their contribution to teaching. Needless to say, the opposite mistake is also frequently made and it consists in accepting the self-defined goals of research units or institutions as the only criteria for their evaluation. It is undeniable, however, that a university department that refuses to train students and spend time doing routine work for outside clients or a technological research center which claims to be doing strictly academic work are possibly examples of misplaced efforts and wasted resources, from the standpoint of their external supporters.

4 - Who should evaluate?

The search for proper criteria is but one aspect of a broader set of concerns: regarding the evaluation of R&D activities. Equally compelling is the question of who is entitled to evaluate and how wide latitude of choice and decisional authority should be granted to evaluators¹⁰.

This question stems from the very nature of R&D efforts. The highly specialized character of scientific and technological work entails a strong measure of self-evaluation and autonomy, since the adequacy of its goals and procedures can hardly be judged by laymen. In addition, scientific researchers normally form a corporate group whose self-esteem and professional pride tend to make them refractory to outside interference. As a result, attempts by planning or regulatory agencies to bring scientific research under close control are liable to fail either because these agencies seldom possess the necessary expertise or because scientists will naturally resist attempts at external control by moving to other jobs. Agencies can, of course, hire scientists into administrative staffs and keep others under control by granting them work privileges and monetary advantages; but the

¹⁰ Stuart S. Blume argues that there is a growing realization that "the right priorities are those which emerge from a properly constituted committee or other forum or process". The same applies to R&D evaluation. See his "Determining Priorities for Science and Technology - A Review of the Unesco Method and its Application". Paper presented to the UNESCO Seminar on Evaluation of Priority Determination Methods in Science and Technology. Paris, 23-30 September. 1983.

best qualified researchers, who tend to be strongly committed to their academic communities and enjoy access to a wider range of professional choice, are not easily attracted¹¹. Clearly, academic commitment is likely to vary across fields of knowledge and different R&D activities, making work in policymaking bodies probably more attractive to, say, scientific researchers devoted to technological development than to those working in less applied areas of research.

An alternative to R&D evaluation from without is the peer review system, widely adopted for the selection of manuscripts submitted for publication in scientific journals and the distribution of research grants through funding agencies and the hiring and promotion of faculty¹². Needless to say, peer review systems are far from full-proofed. But they can easily muster the required expertise and are largely accepted as legitimate by those subject to their evaluation¹³.

A crucial problem of peer review systems is of course the selection of constituent members. Such bodies are not immune to work as representatives of the corporate interests of scientific researchers, particularly when their members are nominated or elected by their own communities. Under such circumstances, peer review bodies are bound to guide their decisions by the interests of local constituencies that do not always coincide with more general criteria of scientific quality or effectiveness. To avoid this pitfall, it is necessary that members of peer judgment bodies be selected as independent and free agents, whose decisional authority rests solely upon their professional competence. It is equally necessary that they be known and respected by their colleagues and hold a public image of expertise and probity if their recommendations are to be at the

¹¹ The notion that academic work entails a strong component of discipline and collegially oriented authority, developed by Burton R. Clark in his analysis of higher education systems applies as well to the scientific research community. See his *the Higher Education System: Academic Organization in Cross-national Perspective* (Berkeley and Los Angeles: University of California Press, 1983, especially Chapter four, "Authority").

¹² See Robert K. Merton and Harriet Zuckerman, "Institutionalized Patterns of Evaluation in Science", in Robert K. Merton, the *Sociology of Science*, edited by Norman W. Storer. Chicago: The University of Chicago Press, 1973. pp. 460-496

¹³ On the limitations of peer review mechanisms, see Stephen Cole, Jonathan R. Cole and Gary A. Simon, "Chance and Consensus in Peer Review", *Science* Vol. 214, N. 20, November 1981. pp. 881-886. See also Rustum Roy, "An Alternative Funding Mechanism", *Science*. Vol. 211, N. 27, March 1981. p. 1377 (editorial) and the ensuing debate in the same journal.

same time consequential and legitimate. The creation of peer review systems is thus a delicate operation, requiring broad consultation and independent judgment. In fact, the process by which such bodies are created is as important as the individual prestige of its members in securing the authoritative nature of its recommendations.

A related problem has to do with the natural inclination of peer review systems to base their evaluation of ongoing activities on previous achievements of research groups and institutions. As a result, they are prone to reinforce the tendency toward concentration of resources and talent within national R&D systems. Such inclination may be curbed or at least partially redressed by means of quota systems that favor certain geographic areas, ethnic groups, types of institutions, and fields of knowledge. The determination of quotas seldom springs from purely technical or scientific criteria, but there is no reason why they should not be established in close consultation with specialists.

5 - Objective vs. subjective indicators of performance

Objective indicators of performance have been often advocated as an alternative that could help reduce some of the biases stemming from the unavoidable subjectivity of peer review mechanisms. There exists a consensus, however, that objective indicators cannot replace qualitative evaluation of scientific procedures and results by qualified reviewers.

Two major types of objective indicators of scientific performance have been developed, one based on cost-benefit considerations and the other on estimates of scientific productivity. Cost-benefit estimates are approximations to efficacy measures. They typically compare the cost of technological innovation with its potential economic benefit as measured by the final product's market value, thus providing a standard for R&D investment decisions.

Cost-benefit analysis has proved to be most useful where expensive and very specific technological development projects are concerned. This approach, however, is of limited value with regard to other areas of scientific development. For one thing, it is practically impossible to estimate eventual spinoffs of R&D programs or future market receptivity to products that are not yet in existence. R&D project costs are also difficult to estimate, to say nothing of external environmental effects that are usually not even taken into account

in many technological development programs¹⁴. Lastly, long-run technological projects deal with future configurations of social and economic needs that are difficult to predict. In such cases, cost-benefit estimates should merely provide general guidelines for decisions have an essentially subjective dimension.

Empirical indicators of scientific productivity, some of which can be rather sophisticated, have also been developed in recent years. Such indicators are measures of effectiveness.

One well-known productivity indicator is the number of publications produced by scientists, institutions, or research groups¹⁵. More elaborate versions, which weight physical production by quality, have also appeared, in particular citation indexes through which the impact of specific scientific contributions on the scientific community can be estimated. Other common indicators include participation in international meetings, academic qualification and awards and the like.

The concern with the development of objective and quantifying measures of scientific performance, however, should not obscure the fact that such indicators dwell, in last instance, on subjective evaluations of specific segments of the scientific community, be it editors of scientific journals academic departments or peer judgment groups. Therefore, however useful they may be as criteria for the evaluation of a whole area of R&D activities, such indicators are certainly inappropriate as guides for the evaluation of specific projects or research groups. "Number of publications," for example, is too rough a measure of scientific merit as its meaning is contingent upon academic discipline and personal style. The meaning of citations can likewise vary and is not rare that articles are cited merely because they happen to report a widely prescribed technicality in lieu of a major scientific contribution. Last but not least, the selection of journals and papers that are to be indexed and included in these measures is always subject to national or linguistic biases that are practically impossible to control.

¹⁴ This is the object of what is known as "technological assessment", constituting a wholly different area of R&D evaluation.

¹⁵ This is part of the growing field of "scientometrics", originally associated with the work of Derek J. de Sola Price. See his *Little Science, Big Science* (New York: Columbia University Press, 1963).

Objective indicators, then, seem to serve two main purposes. On the one hand, they are useful for the evaluation of long-term developments and may help detect R&D efforts that have lagged behind or that are of exceptional quality. On the other hand, objective measures of scientific performance can provide review committees with useful points of reference for making the comparisons that form the basis of peer evaluation.

6 - The uses of R&D evaluation

Research and development evaluation can serve important purposes. For planning and science policy agencies, it can provide operational guidelines that are necessary if limited resources are to be properly distributed to a large number of applicants. At higher levels of authority, it can assure that resources are being applied according to the institution's goals. In general, it can lead to better utilization of existing resources and contribute to the improvement of scientific and technological work.

R&D evaluation may also have other less direct but perhaps more important consequences. It can bestow legitimacy upon national R&D systems, enhance their national and international prestige, and assure access to needed resources. Moreover, an established system of evaluation extensively based on peer review mechanisms provides scientific researchers with a sense of efficacy and participation in the conduct of their own affairs. The establishment of legitimate standards of evaluation can also have important educational effects over national R&D systems. By providing criteria for the ranking of research groups and institutions along a gradient of scientific excellence, and by establishing the associated system of rewards, R&D evaluation mechanisms help create and crystallize models of scientific performance.

Therefore, far from being a mere technical device for resource distributions evaluation can play an important role in the improvement and development of national R&D systems. In reality, it may well be that an institutionalized and respected system of R&D evaluation constitute a crucial condition for the sustained support for science and technology in modern societies.